ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 6 - Due date 03/16/22

Tatiana Sokolova

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

Once you have the project open the first thing you will do is change "Student Name" on line 3 with your name. Then you will start working through the assignment by **creating code and output** that answer each question. Be sure to use this assignment document. Your report should contain the answer to each question and any plots/tables you obtained (when applicable).

When you have completed the assignment, **Knit** the text and code into a single PDF file. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A06_Sp22.Rmd"). Submit this pdf using Sakai.

Questions

This assignment has general questions about ARIMA Models.

Packages needed for this assignment: "forecast","tseries". Do not forget to load them before running your script, since they are NOT default packages.

```
#Load/install required package here
library(forecast)
library(tseries)
library(sarima)
```

```
## Warning: package 'sarima' was built under R version 4.1.3
```

Q1

Describe the important characteristics of the sample autocorrelation function (ACF) plot and the partial sample autocorrelation function (PACF) plot for the following models:

(a) AR(2)

> Answer: For the autoregressive process, the ACF would tail off, showing a slow decay. The PACF, on the other hand, would cut off. Since the auto-regressive process is forecasting two steps ahead of time (two-period forecast), the PACF would cut off at lag 2.

(b) MA(1)

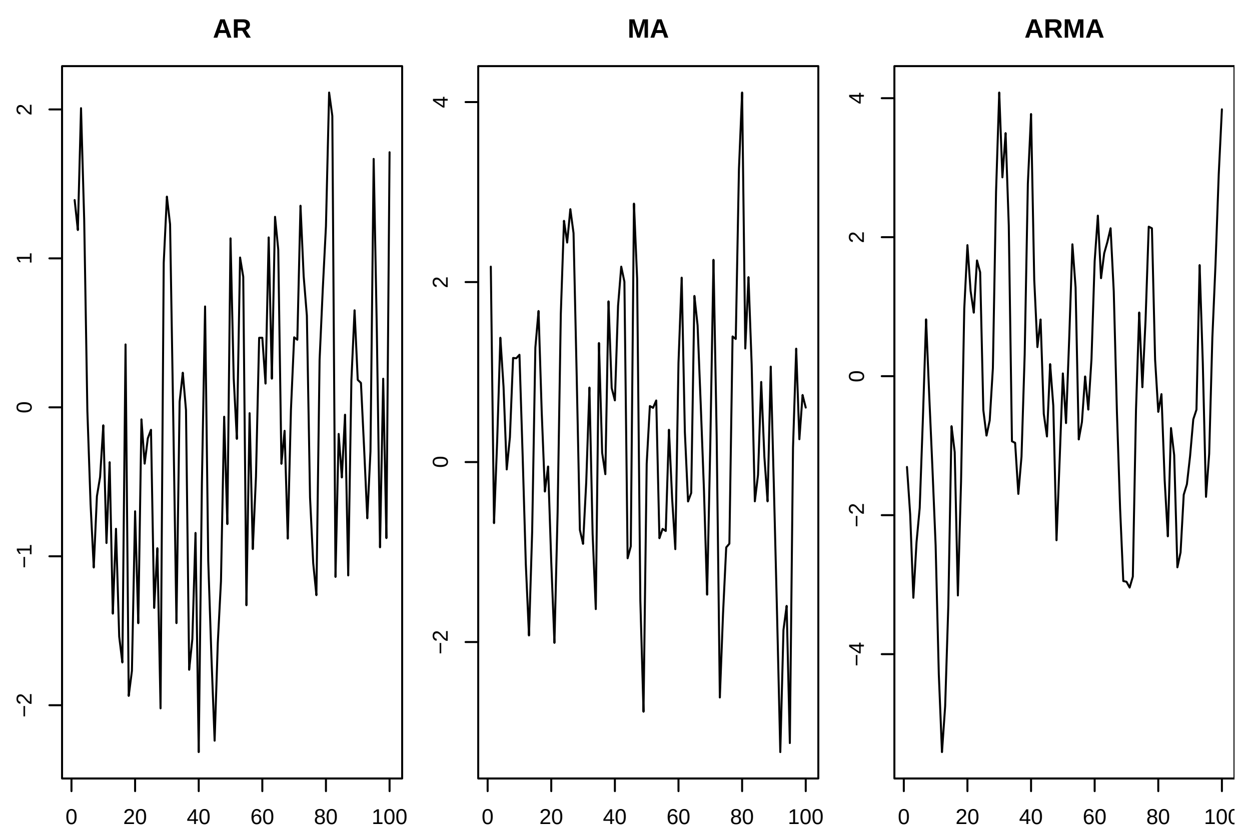
> Answer: For the moving average process, the ACF would cut off. The PACF would then tail off, or decay slowly. Since the moving average process is forecasting only 1 step ahead of time (one-period forecast), the ACF would cut off at lag 1.

Q2

Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference the series, we don't need to specify the "I" part and we can use the short version, i.e., the ARMA(p, q). Consider three models: ARMA(1,0), ARMA(0,1) and ARMA(1,1) with parameters $\phi = 0.6$ and $\theta = 0.9$. The ϕ refers to the AR coefficient and the θ refers to the MA coefficient. Use R to generate n=100 observations from each of these three models

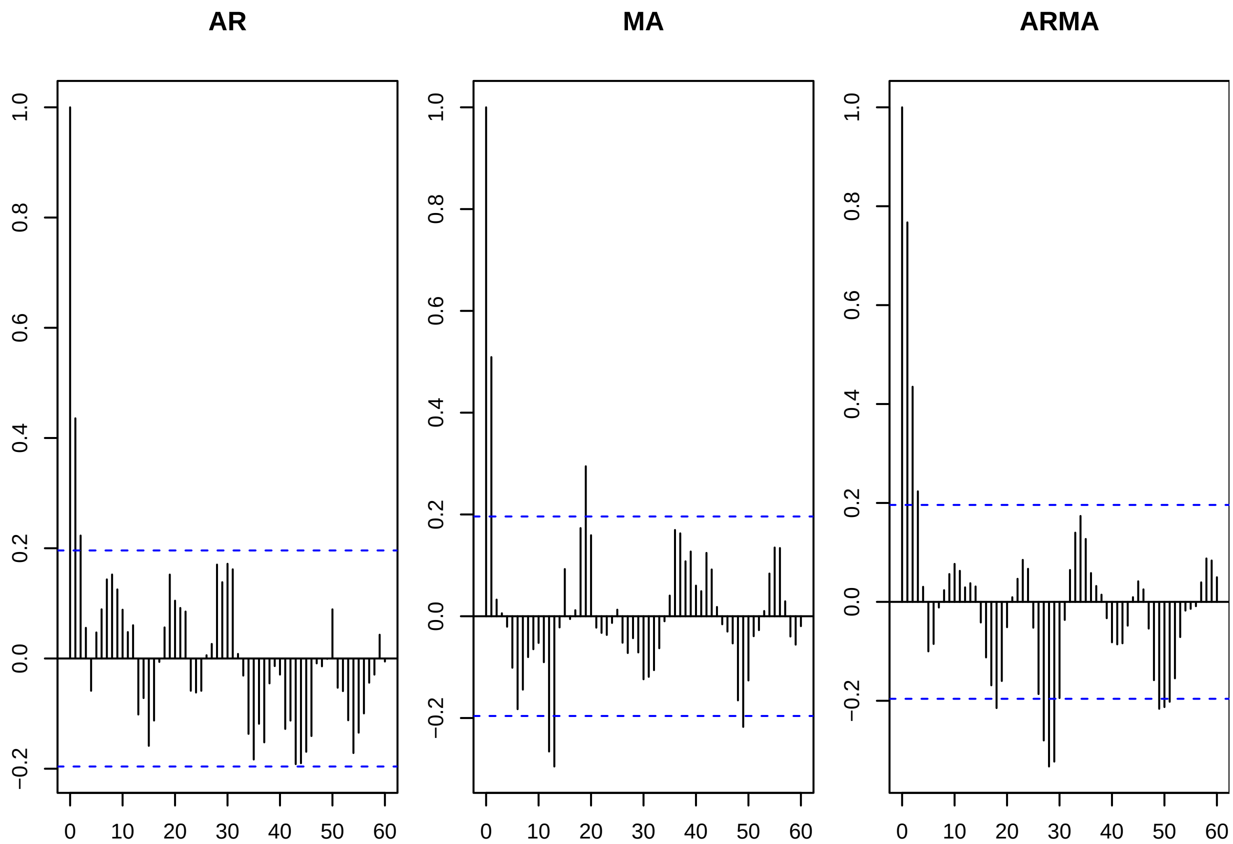
```
obs=100
AR <- arima.sim(list(order= c(1,0,0), ar = 0.6), n=obs)
MA <- arima.sim(list(order= c(0,0,1), ma = 0.9), n=obs)
ARMA <- arima.sim(list(order= c(1,0,1), ar = 0.6, ma = 0.9), n=obs)

#confirming proper creation
par(mar=c(3,3,3,0));par(mfrow=c(1,3))
plot(AR, main = "AR", xlab="Observation")
plot(MA, main = "MA", xlab="Observation")
plot(ARMA, main = "ARMA", xlab="Observation")
```



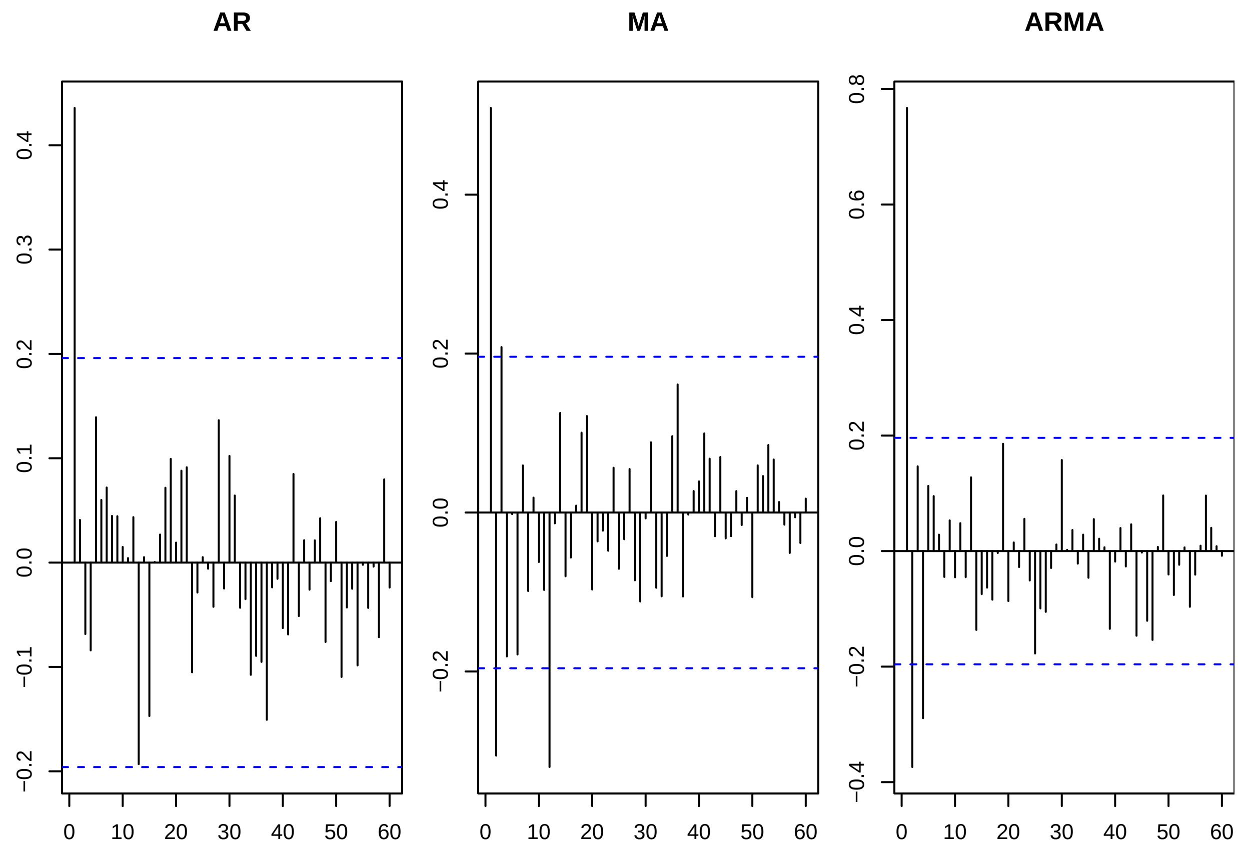
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command par(mfrow=c(1,3)) that divides the plotting window in three columns).

```r
par(mar=c(3,3,3,0));par(mfrow=c(1,3))
acf(AR,lag.max=60,main="AR")
acf(MA,lag.max=60,main="MA")
acf(ARMA,lag.max=60,main="ARMA")
```



(b) Plot the sample PACF for each of these models in one window to facilitate comparison.

```r
par(mar=c(3,3,3,0));par(mfrow=c(1,3))
pacf(AR,lag.max=60,main="AR")
pacf(MA,lag.max=60,main="MA")
pacf(ARMA,lag.max=60,main="ARMA")
```



(c) Look at the ACFs and PACFs. Imagine you had these plots for a data set and you were asked to identify the model, i.e., is it AR, MA or ARMA and the order of each component. Would you be able to identify them correctly? Explain your answer.

Answer: *Note: Answers below are in response to previously simulated models and when knit, the plots may not reflect the observations listed below.*

For the first plot, I could determine that it is AR(1) using the PACF because of the significant cut off after lag 1. Using the ACF alone would not be helpful as the graph does not have dramatic decay and could be mistaken for an ARMA model.

For the second plot, I could determine that it is MA(1) from the ACF because it has a significant cut off at lag 1. I could not use the PACF alone to determine if this is an MA(1) because it only has a slight decay.

For the third plot, using both the ACF and PACF would be helpful in determining that it is an ARMA model as both show a geometric decay, though the decay in the ACF is more prominent.

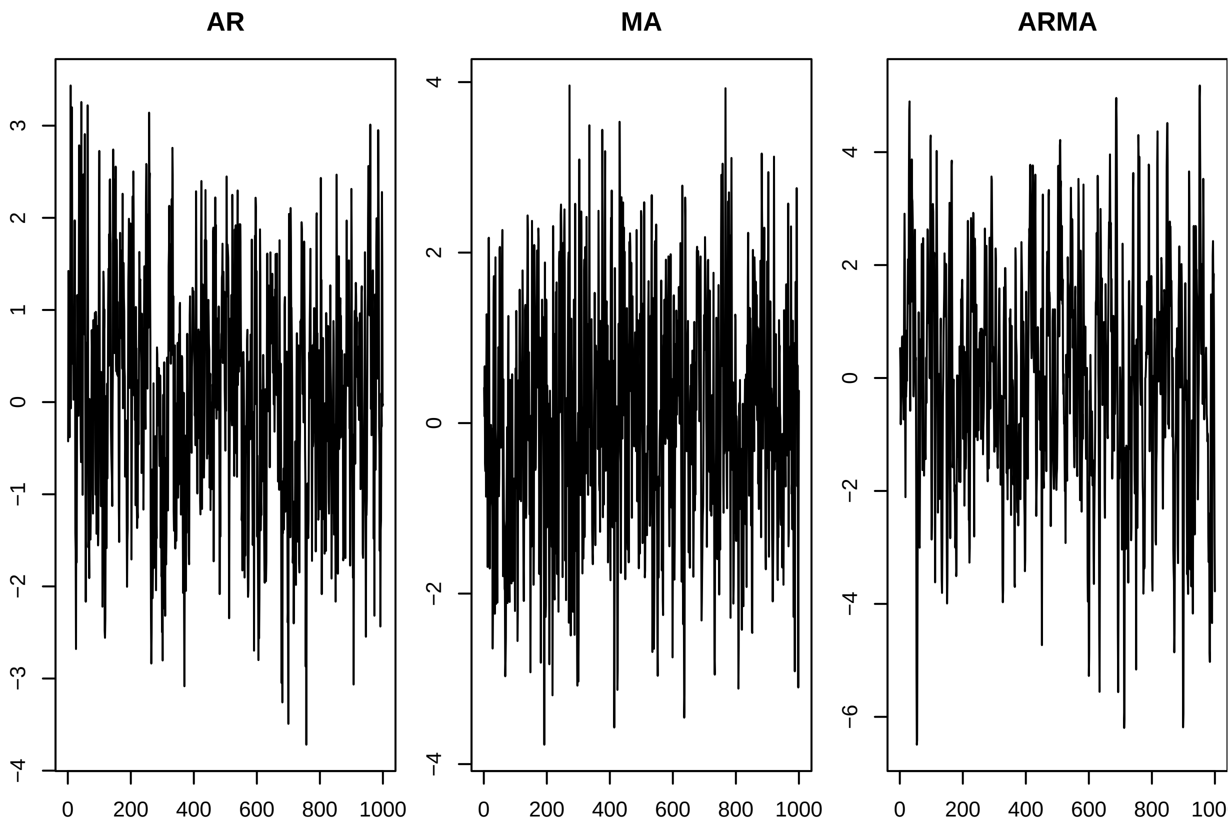
(d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

Answer: The PACF for the AR model should theoretically match the phi value of 0.6 at lag 1. However, the PACF appears to show lag 1 at around 0.5. This is likely due to the small sample size for the simulation. It is not possible to gauge the theta value from the MA model's PACF or ACF (and therefore from the ARMA model as wel) as the moving average is based on previous errors and not the correlation of past values.

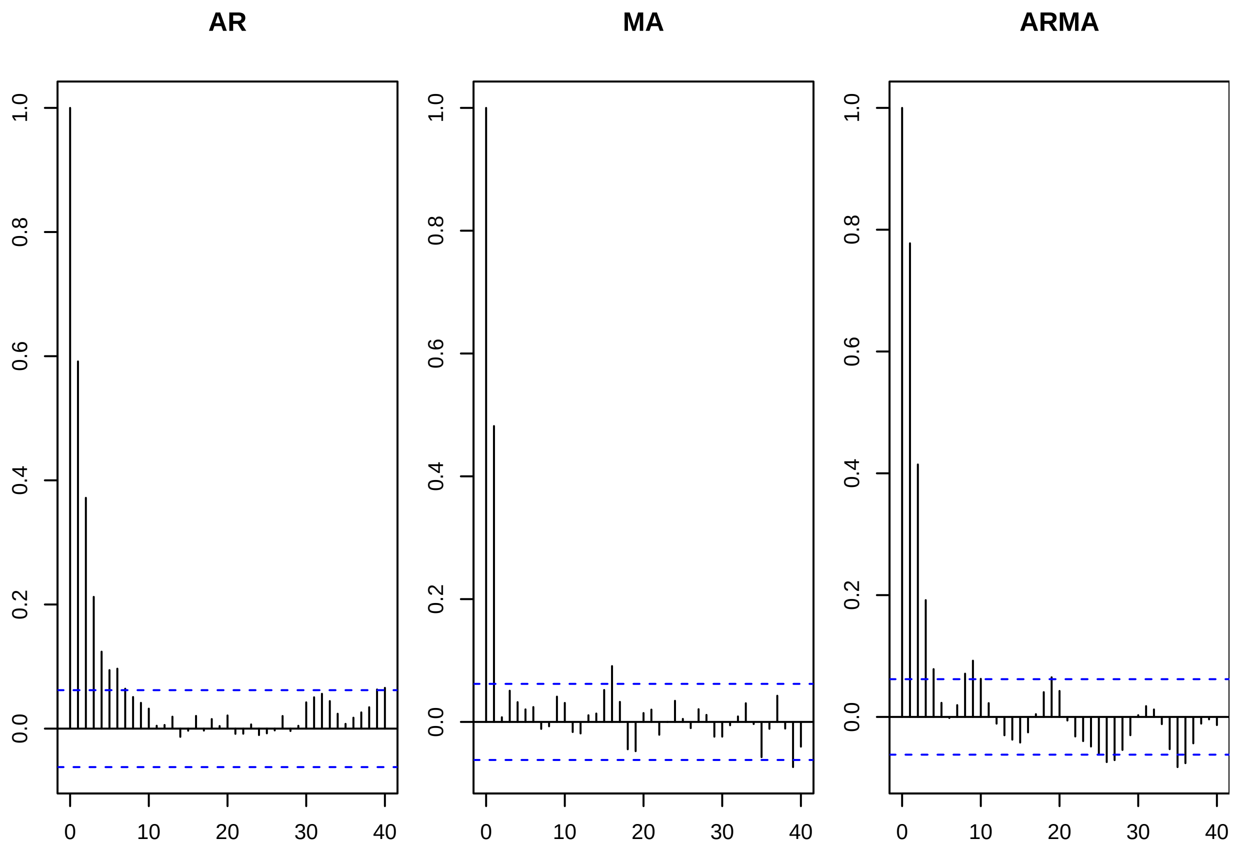
(e) Increase number of observations to n=1000 and repeat parts (a)-(d).

```
obs2=1000
AR2 <- arima.sim(list(order= c(1,0,0), ar = 0.6), n=obs2)
MA2 <- arima.sim(list(order= c(0,0,1), ma = 0.9), n=obs2)
ARMA2 <- arima.sim(list(order= c(1,0,1), ar = 0.6, ma = 0.9), n=obs2)
```

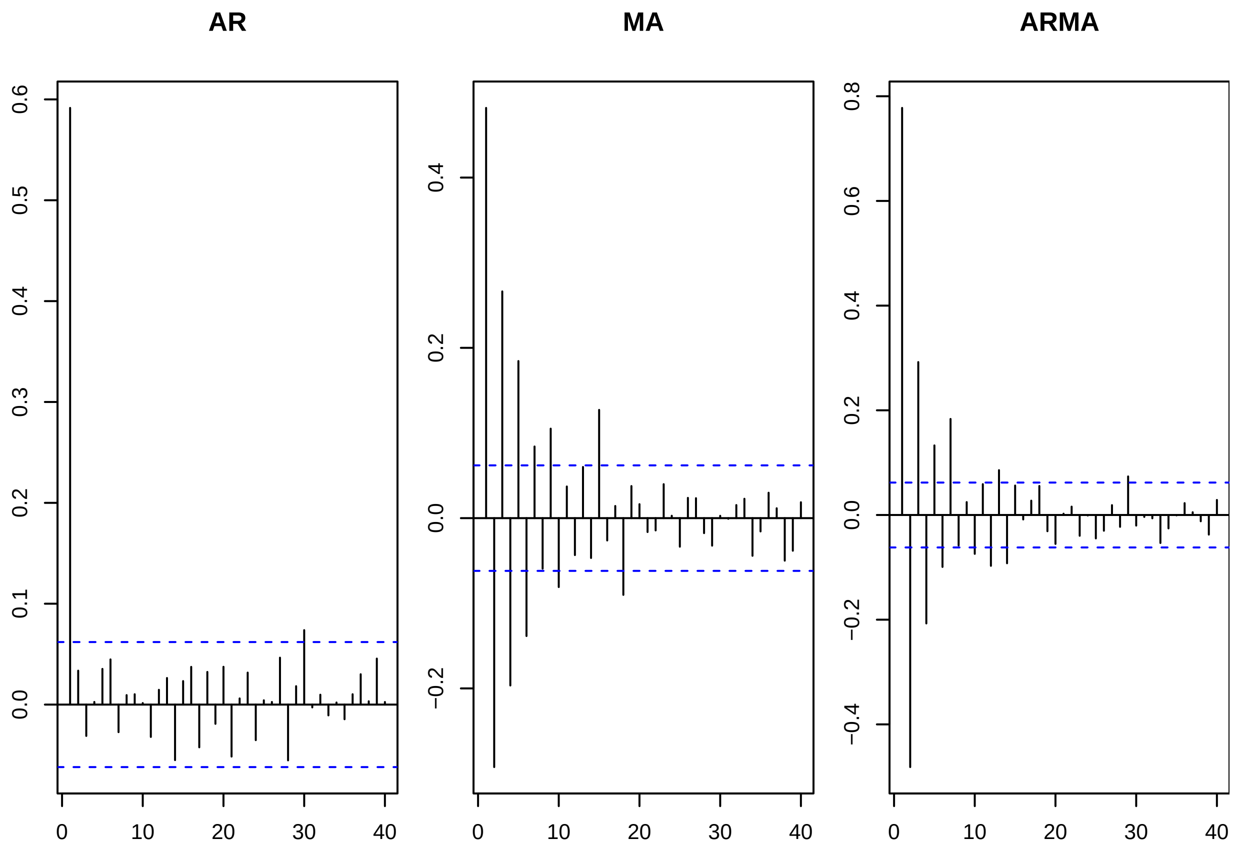
```
#confirming proper creation
par(mar=c(3,3,3,0));par(mfrow=c(1,3))
plot(AR2, main = "AR", xlab="Observation")
plot(MA2, main = "MA", xlab="Observation")
plot(ARMA2, main = "ARMA", xlab="Observation")
```




```
#(e-a) ACF
par(mar=c(3,3,3,0));par(mfrow=c(1,3))
acf(AR2,lag.max=40,main="AR")
acf(MA2,lag.max=40,main="MA")
acf(ARMA2,lag.max=40,main="ARMA")
```




```r
#(e-b) PACF
par(mar=c(3,3,3,0));par(mfrow=c(1,3))
pacf(AR2,lag.max=40,main="AR")
pacf(MA2,lag.max=40,main="MA")
pacf(ARMA2,lag.max=40,main="ARMA")
```



(e-c) > Answer: For the first plot, I could determine that it is AR(1) using the PACF because of the significant cut off after lag 1. With the increase in observations, it is also more clear from the ACF that the first plot is AR(1) as it shows significant decay.

For the second plot, I could determine that it is MA(1) from the ACF because it has a significant cut off at lag 1. With the increase in observations, I can now use the PACF as well to check that this plot is MA(1) because it has a now significant geometric decay (but could not solely use the PACF as it looks similar to that of an ARMA model).

For the third plot, using both the ACF and PACF would be necessary in determining that it is an ARMA model as both show a significant decay.

(e-d) > Answer: The PACF for the AR model now matches the phi value of 0.6 with the increase in observations. The theta value still cannot be identified from the MA or ARMA model.

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

(a) Identify the model using the notation ARIMA$(p, d, q)(P, D, Q)_s$, i.e., identify the integers p, d, q, P, D, Q, s (if possible) from the equation.

> Answer: Notation below:

SARIMA$(1, 0, 1)(1, 0, 0)_{12}$

(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

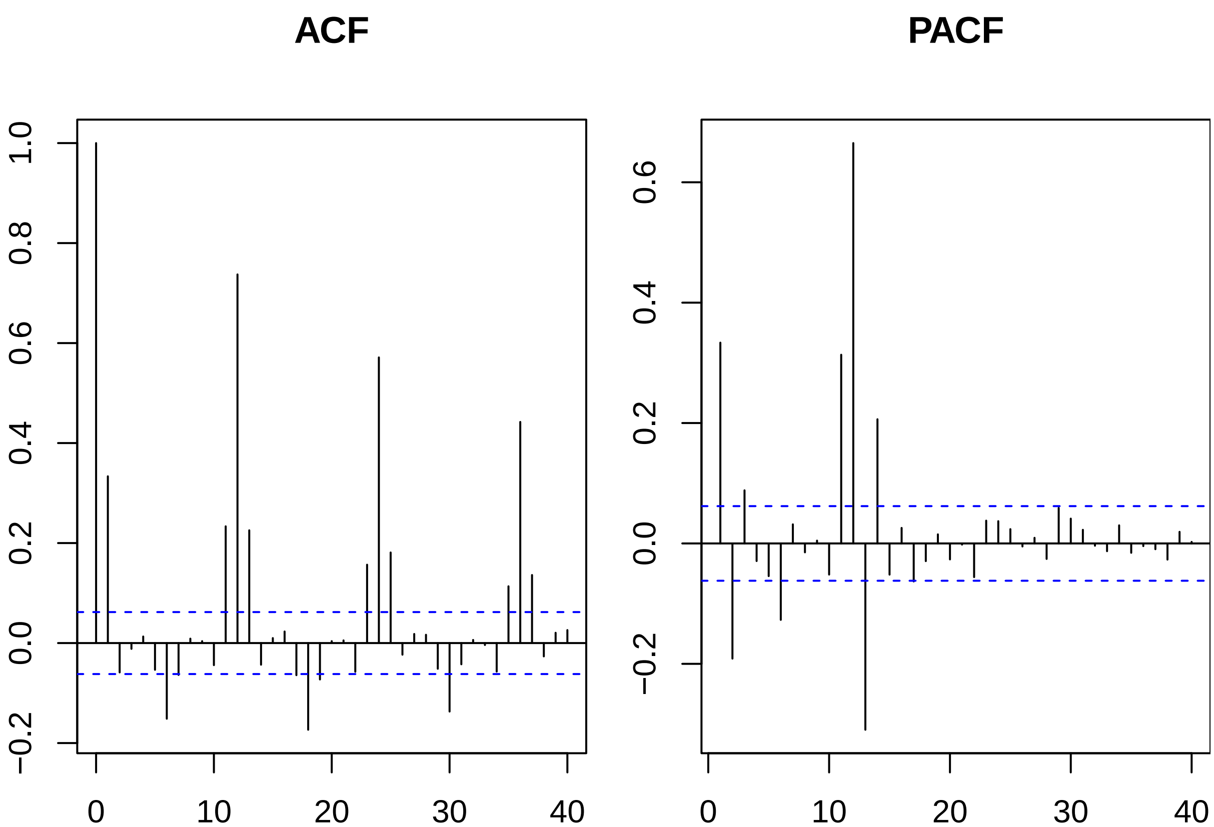
> Answer: Coefficients listed below:

$\phi_1 = 0.7$, $\phi_{12} = 0.25$, $\theta_1 = 0.1$

Q4

Plot the ACF and PACF of a seasonal ARIMA$(0,1) \times (1,0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using R. The 12 after the bracket tells you that $s = 12$, i.e., the seasonal lag is 12, suggesting monthly data whose behavior is repeated every 12 months. You can generate as many observations as you like. Note the Integrated part was omitted. It means the series do not need differencing, therefore $d = D = 0$. Plot ACF and PACF for the simulated data. Comment if the plots are well representing the model you simulated, i.e., would you be able to identify the order of both non-seasonal and seasonal components from the plots? Explain.

```
sarima_model <- sim_sarima(model=list(ma=0.5,sar=0.8, nseasons=12), n=1000)
par(mar=c(3,3,3,0));par(mfrow=c(1,2))
acf(sarima_model,lag.max=40,main="ACF")
pacf(sarima_model,lag.max=40,main="PACF")
```



I think these plots represent the simulated model well. The ACF shows that this is a MA(1) model because of the cut off at lag 1. However, the multiple spikes at the seasonal lag (12, 24, 36, etc.) in the ACF, imply that it has a seasonal AR component. The PACF confirms that it is a MA(1) model due to the slow decay until the seasonal lag. However, the single spike at lag 12 implies that it has a seasonal AR component as well.